POD Tutorial Part I
Dealing with Hit/Miss Data

William Q. Meeker

wqmeeker@iastate.edu

Center for Nondestructive Evaluation

Department of Statistics

Iowa State University

Empirical POD
Broad Overview

- POD from Hit-Miss Data
- POD ahat versus a
- Advanced Topics in Empirical POD

Dealing with Hit/Miss Data Overview

- What is POD?
- What is PFA?
- Sources of data
- Hit/Miss data
- Binary regression model for Hit/Miss data
- POD from the binary regression model

What is POD???

- Probability of Detection
- Fraction of flaws, in a large number of inspections, that will be detected
- May be (and typically is) conditional on flaw characteristics (e.g., size and depth)
- By making model assumptions, can be expressed as part of a probability or statistical model
- POD will depend on the detection rule (threshold)

What is PFA???

- Probability of false alarm
- Fraction of inspection opportunities, in a large number of inspections, that will call a "detect" when no flaw is present
- Will depend on the noise level and threshold
- Definition of "inspection opportunity" is important (what is the denominator?). Consider the extremes:
  - per forging
  - per pulse on a forging UT inspection?

POD Studies
(Sources of POD Data)

- Designed experiments
- Field finds

Designed Experiments

- Most common form of POD data (in my experience)
- Need a collection of representative flaws. Possible choices:
  - Real
  - Synthetic simple (flat bottom hole, notch, synthetic hard alpha)
  - Synthetic realistic (fatigue cracks)
- Conduct inspection experiment on the specimens

Field-Find Data

- Typically only a limited number of field finds in most applications
- Need careful flaw characterization (e.g., sizing)
- Must use special statistical methods (truncated regression) to account for misses

Data Giving Hit/Miss versus Flaw Size

- For each inspection opportunity on a unit with a known flaw, we have a detect (Hit) or not (Miss)
- Model POD as a function of flaw size
- Frequently used in FPI POD estimation where good quantitative data is not available

Sonic IR Hit/Miss Data

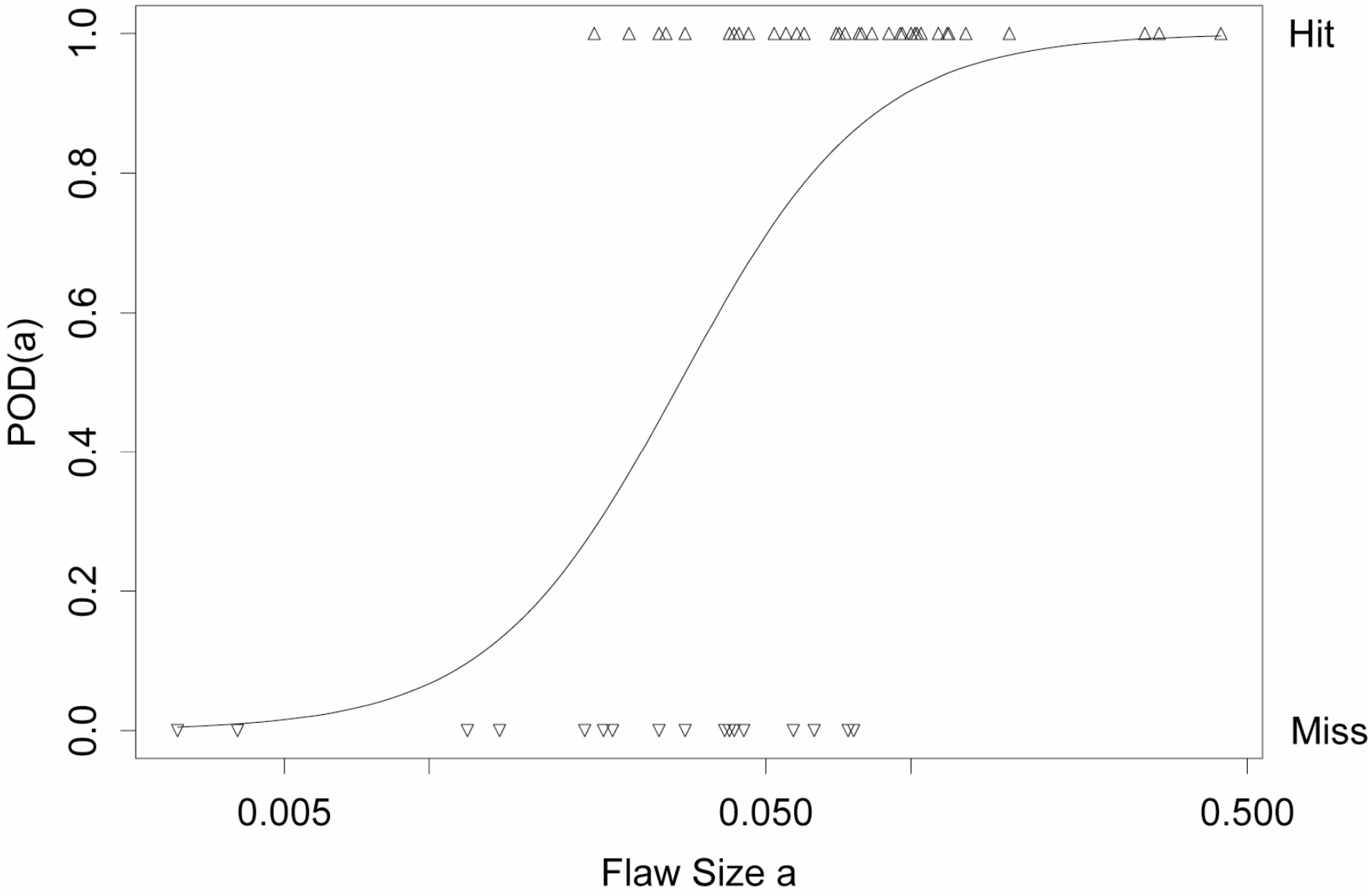
Sample number	Detectable?	Crack Length
1	0	0.0125
2	1	0.032
3	0	0.0075
4	0	0.016
5	1	0.025
6	0	0.0096
7	0	0.0215
8	0	0
9	1	0.038
10	0	0.0125
11	0	0.0085
12	1	0.021
13	1	0.028

68	1	0.036
69	1	0.0205
70	0	0.012

Estimation of POD from Hit\Miss Data

- Model POD as a function of flaw size (or other characteristic(s))
- The appropriate statistical technique is “binary regression.”
- Common models used in binary regression are the “logistic model” and the “probit model.”

POD Estimate from FPI Hit/Miss Data



Binary Regression Model

- Pr(Hit) is a function of flaw size. Commonly:

$$POD(a) = \Pr(\text{Hit}) = \Phi\left(\frac{\log(a) - \mu}{\sigma}\right) = \Phi(\beta_0 + \beta_1 \log(a))$$

- Use Maximum likelihood (ML) to estimate the parameters $\beta_0 = -\mu/\sigma$ and $\beta_1 = 1/\sigma$.
  - For logit $\Phi(z) = \exp(z)/(1+\exp(z))$
  - For probit $\Phi(z)$ is the standard normal cumulative distribution function.

MINITAB Estimation of Binary Regression Model Parameters

- The MINITAB "Logistic regression" procedure will estimate the parameters β_0 and β_1.
- After the parameters have been estimated, POD (and possibly confidence intervals) could be obtained by
  - Writing a MINITAB macro
  - Importing results to Excel and writing a macro or procedure there
  - Using other simple software (e.g. MATLAB).

MINITAB Worksheet Sonic IR Hit/Miss Data (blank specimens removed)

Sample number	Detected	CrackLength	Log10CrackLength
1	0	0.0125	-1.90309
2	1	0.0320	-1.49485
3	0	0.0075	-2.12494
4	0	0.0160	-1.79588
5	1	0.0250	-1.60206
6	0	0.0096	-2.01773
7	0	0.0215	-1.66756
9	1	0.0380	-1.42022
67	0	0.0125	-1.90309
68	1	0.0360	-1.44370
69	1	0.0205	-1.68825
70	0	0.0120	-1.92082

MINITAB Binary Regression Output

Binary Logistic Regression: Detected versus Log10CrackLength

Link Function: Logit

Response Information

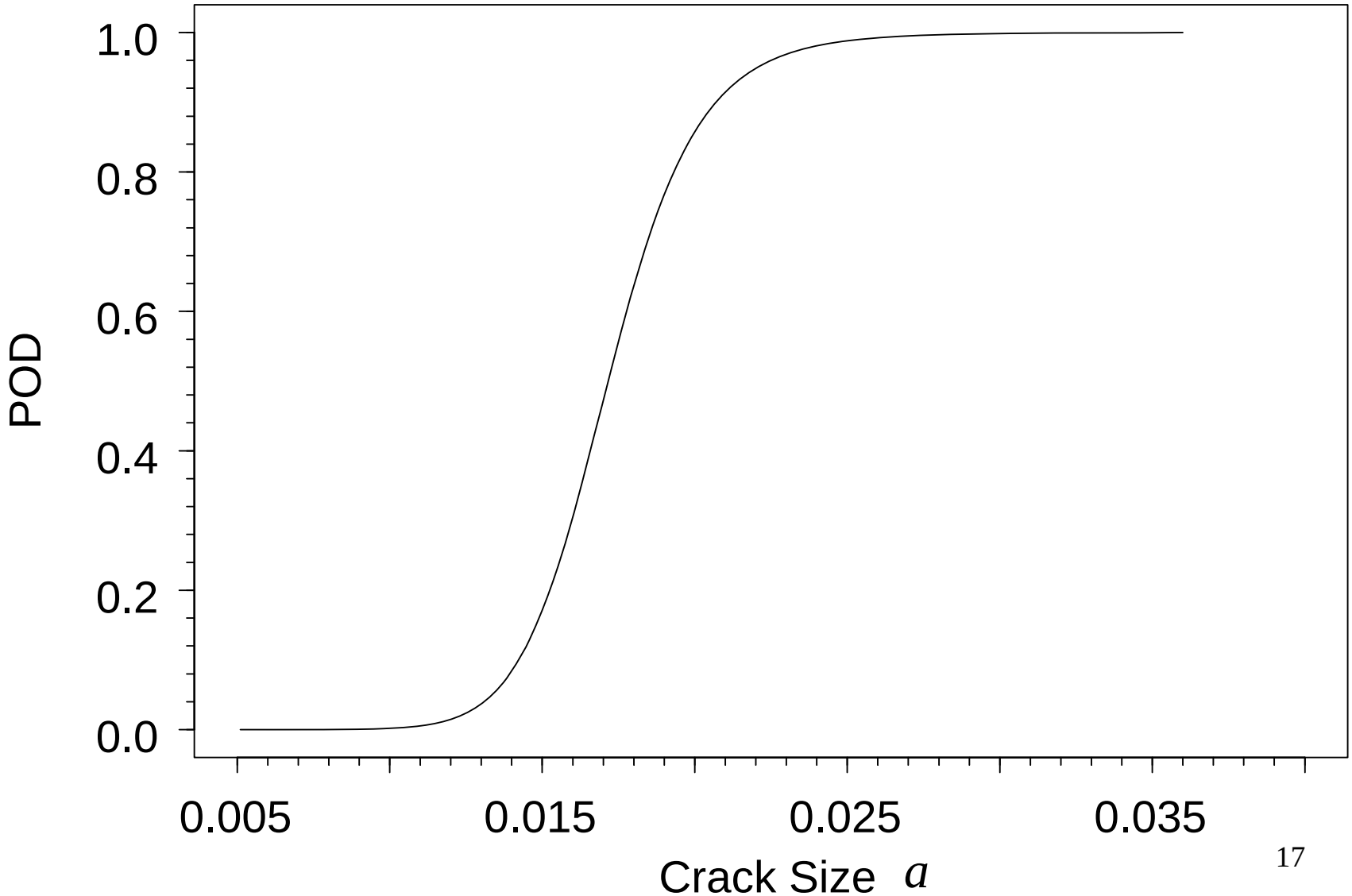
Variable	Value	Count	
Detected	1	30	(Event)
	0	30	
	Total	60	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower
Constant	**47.7243**	14.2260	3.35	0.001		
Log10CrackLength	**27.0323**	8.07139	3.35	0.001	5.49512E+11	74040.28

$$\hat{POD}(a) = \Phi(\hat{\beta}_0 + \hat{\beta}_1 \log_{10}(a)) = \Phi(47.72 + 27.03 \times \log_{10}(a))$$

$$\hat{POD}(a) = \Phi(\hat{\beta}_0 + \hat{\beta}_1 \log_{10}(a)) = \Phi(47.72 + 27.03 \times \log_{10}(a))$$



Logistic Model POD 95% Confidence Limits (pointwise and simultaneous)